

Psychophysical Harmony: A New Argument for Theism

1. Introduction

This paper develops a new argument from consciousness to theism: the *argument from psychophysical harmony*. Roughly, psychophysical harmony consists in the fact that phenomenal states are correlated with physical states and with one another in strikingly fortunate ways. For example, phenomenal states are correlated with behavior and functioning that is *justified* or *rationalized* by those very phenomenal states (e.g., pain is correlated with avoidance behavior), and phenomenal states are correlated with verbal reports and judgments that are made true by those very phenomenal states (e.g., we say, “I’m in such-and-such phenomenal state,” and sure enough, we are).

In §2, we argue that psychophysical harmony is strong evidence for theism. Since God has reason to design the psychophysical laws in order to bring about the values realized by psychophysical harmony, theism makes harmony much more likely than it would otherwise be. In our initial presentation of the argument, we rely on two controversial metaphysical assumptions: *dualism* (roughly, the view that the phenomenal and physical domains are ontologically distinct and co-fundamental) and *the causal completeness of the physical* (roughly, the view that every physical event involved in human behavior and brain functioning has a sufficient causal explanation in terms of prior physical events). While these assumptions are convenient for an initial presentation of the argument, we show in §3 that they are ultimately inessential: the argument still works if we accept (or we are open to) alternative views about consciousness, such as dualist interactionism, physicalism, idealism, or Russellian monism. In §4, we draw a parallel between our argument and the argument from cosmological fine-tuning

and highlight a potential advantage of our argument, namely, that our argument is not threatened by multiverse hypotheses.

For simplicity's sake, we frame this as an argument for theism, which we can understand as the claim that the universe was created by an all-powerful, all-knowing, and all-good being. But the data to which we appeal may be equally good evidence for other hypotheses. The most obvious candidates are other views on which the universe is somehow ordered towards realizing the values found in harmoniously conscious beings. Such views might include John Leslie's axiarchism (2001) and other broadly Platonic views, Thomas Nagel's (2012) view that the universe is intrinsically teleologically ordered toward the realization of value, Paul Draper's (2017) "aesthetic deism," or other views on which the creator is not all-powerful, all-knowing, or all-good. We will not argue here that theism is to be preferred to these "theism-adjacent" views; that choice may need to be made on other grounds. Relatedly, our target can be understood as, not exactly atheism as such, but atheism in its standard "naturalistic" form—that is, atheism together with the claim, very roughly, that the universe is purposeless, not teleologically ordered toward the realization of any kind of value, whether extrinsically (e.g., by God or a Platonic Form) or intrinsically.

Several other philosophers have presented arguments from consciousness to theism (Adams 1992; Swinburne 2004, ch. 9; Moreland 2008; such arguments trace back at least to Locke 1689/1996: Bk IV, ch. X, §10). These arguments tend to share the same general structure. First, they argue on independent grounds against physicalist accounts of consciousness. Second, they argue that, given the failure of these accounts, any explanation of the correlations between conscious states and physical states will need to be a personal explanation (i.e., an explanation in terms of an agent's intentional action) rather than a scientific one. Third, they argue that there is,

or probably is, an explanation for these correlations. Possible justification for this includes the general idea that it's theoretically virtuous to explain as much as possible, as well as the more specific idea that, as Swinburne argues, the psychophysical correlations are so complicated that it's very intrinsically improbable that they would obtain as brute facts. From this it follows that there is, or probably is, a personal explanation for psychophysical correlations. Finally, they argue that the best candidate for the agent referenced by this personal explanation is God.

Our argument differs from these in at least three crucial respects. The first is our argument's main explanandum. We appeal, not just to the existence of psychophysical correlations (or complex psychophysical correlations), but specifically to psychophysical harmony. The success of these other arguments may therefore be to some extent compatible with our argument succeeding and having independent force. A second (related) difference is that, unlike the arguments of Swinburne and Adams, ours does not rely on the premise that the basic psychophysical laws are extremely complex, taking the form of an enormous list of specific causal connections between brain states and phenomenal states which can't be derived from any reasonably simple set of underlying psychophysical principles. Such claims about the degree of complexity in the ultimate psychophysical laws seem to us highly speculative, so it is an advantage that we can remain neutral on such claims. Moreover, even if the basic psychophysical laws *are* enormously complex, this in itself may not be strong evidence for theism apart from (something like) psychophysical harmony. Imagine we discover that the basic laws governing X particles are extremely complex, taking the form of an enormous list. If these laws give rise to X-particle behavior that is no more special or valuable than the behavior we would get under other possible laws, this discovery would not be strong evidence for theism. While any specific, complicated set of laws is very improbable on atheism, such laws will not be any more probable

on theism, unless there is something valuable about those laws in comparison with alternative laws (cf. Oppy 2006, 400). On the other hand, if the basic laws governing X particles gave rise to X-particle behavior that is distinctly valuable in comparison with what we would get under alternative laws, this plausibly *would* be evidence for theism. Likewise, complexity in the psychophysical laws is not strong evidence for theism apart from the claim that those laws are somehow distinctly valuable in comparison with alternative psychophysical laws--perhaps because they realize psychophysical harmony.¹

Third, our argument doesn't require independently refuting physicalism. More specifically, as we explain in §3, psychophysical harmony may itself be evidence against physicalism, but we need not rely on any of the “standard” anti-physicalist arguments. In fact, we can proceed from premises granted by most physicalists. This should give our argument substantially wider appeal.²

2. The Argument from Psychophysical Harmony for Theism

2.1 Psychophysical harmony

Here we present the argument from psychophysical harmony in its basic form. We'll initially make some substantive assumptions about the metaphysics of consciousness. In §3, we show that the assumptions can be relaxed without seriously affecting the argument. The assumptions are:

¹ Swinburne himself (2004, 124-131) suggests that God would want to create conscious *embodied* creatures because being embodied plays an important role in our exercising agency, acquiring knowledge, etc. Our arguments later will suggest that our agency, acquisition of knowledge, etc. would be seriously hindered if not for psychophysical harmony. So to some extent, our invocation of psychophysical harmony might be seen as complementary to Swinburne's account.

² Ben Page (forthcoming) offers a version of the argument from consciousness which presupposes materialism. The idea is that the complex material arrangements needed for consciousness are more likely to arise on theism than atheism. In a way, this makes his argument more similar to traditional arguments from cosmological fine-tuning or biological design. Again, the explanandum of our argument is quite different.

Dualism: Phenomenal truths and physical truths are distinct and co-fundamental, with neither class of truths grounded in the other. Physical and phenomenal states are linked by metaphysically contingent fundamental laws of nature that specify which physical configurations give rise to consciousness in its various forms.

Causal completeness: Every physical event involved in human behavior and brain functioning has a sufficient causal explanation in terms of prior physical occurrences.

Together, these assumptions imply the disjunction of epiphenomenalist and overdeterminationist dualism.

Given dualism, we think that the very existence of consciousness is at least some evidence for theism. If consciousness is ontologically distinct from any physical properties, a physical universe can host consciousness only by adding it to its supply of fundamental features. That it would do so is unsurprising if our universe was designed by a being which aims to realize value. A world with intricate arrangements of matter but no experience is clearly missing some important kinds of value, and perhaps missing value altogether. It's far more surprising that consciousness should exist (and that there should be fundamental laws governing its occurrence) if the universe is not ordered in any way toward the realization of value. There would be no reason to expect it to exist, and the fundamental laws would be simpler if it didn't. (By contrast, that mass or charge--properties without apparent normative significance--are included among the universe's basic features doesn't seem much more surprising on atheism than theism.)

However, the main focus of our argument will be a different set of facts about consciousness, which we'll call the facts of "psychophysical harmony":

Psychophysical Harmony: States of consciousness are related to each other, and to physical states, in strikingly harmonious ways—ways that seem extremely *lucky*, or involve many striking apparent coincidences.

Various instances of psychophysical harmony have been emphasized in recent work by Adam Pautz (2020), David Chalmers (2018), Philip Goff (2018), Hedda Hassel Mørch (2017, 2020), Harold Langsam (2011), Noa Latham (2000), and Bradford Saad (2019). We'll discuss two main types of harmony: *normative harmony* and *semantic harmony*.

2.2. Normative harmony

Many examples of psychophysical harmony are cases of what Adam Pautz (2020: 5) calls *normative harmony*, which he defines as follows:

Normative harmony: In every case, the psychophysical laws correlate a physical functional state P with a distinct conscious experience C whose essential normative role in providing reasons is harmonious with the causal role of P in generating verbal and other responses.

This will be easiest to explain with some examples (which closely mirror some of Pautz's own examples). Some are examples of *hedonic* harmony; others are examples of *epistemic* harmony. We'll also discuss a third, somewhat overlapping type of normative harmony, *cognitive* harmony, but will rely on it less because it requires a controversial background assumption.

Here's a hedonic example. A damaging stimulus causes physical state X, a certain biochemical or computational state of your brain. X causes you to avoid or eliminate the stimulus in the future. Conveniently, the psychophysical laws map X onto the experience of pain, an intrinsically bad experience which essentially provides one with reason to avoid or eliminate it. So the psychophysical laws correlate X with a phenomenal state whose essential *normative* role

harmonizes with the *functional* role of X. And this isn't a random fluke, but a pretty general truth about relevantly similar functional states and their associated hedonic states: we systematically avoid unpleasant experiences and pursue pleasant experiences. This is *hedonic harmony*.³ To be clear, the surprising fact here is not that there is a physical state that plays the pain role (tracking bodily damage, producing avoidance behavior, and so forth). Presumably this fact has a straightforward evolutionary explanation. What's surprising is that the actual psychophysical laws map it onto an experience whose essential normative role harmonizes with this functional role. Since evolutionary forces cannot affect the psychophysical laws, it's hard to see how an evolutionary explanation of this harmonious correspondence would even get off the ground. (Note that we are not rejecting the standard evolutionary explanation for why we feel pain in response to harmful stimuli. *Given* that pain is lawfully linked to avoidance behavior and the like, it makes perfect evolutionary sense that we would experience pain in response to harmful stimuli. But this evolutionary explanation *presupposes* normative harmony; it does not explain it. There is nothing inappropriate about this presupposition when we are doing evolutionary biology; it is not the evolutionary biologist's job to explain the character of the psychophysical laws.)

Hedonic harmony seems very lucky. The psychophysical laws could conceivably have mapped X onto pleasure, while mapping the actual neural basis of pleasure onto pain. In this pleasure/pain inversion scenario, we would systematically avoid a state we have reason to pursue (pleasure), and systematically pursue a state we have reason to avoid (pain). Our lives would be a pathetic farce: we would cower from pleasurable experiences and happily inflict pain on our loved ones. Alternatively, the psychophysical laws could have correlated X with some

³ There may be intrinsically good or bad phenomenal experiences besides pleasures and pains (perhaps, say, certain aesthetic experiences are like this). If so, they seem to exhibit the same kind of harmony. We would count these as part of hedonic harmony, too, even though "hedonic" is not quite the right word for them.

evaluatively neutral state, resulting in a less extreme mismatch. Either way, our behavior and functioning would be wildly out of line with the behavior and functioning that is *justified* or *rationalized* by our phenomenal states.

A second type of normative harmony is the harmonious match between the *epistemic* reasons provided by various phenomenal states and their associated functional states. Call this *epistemic harmony*. You undergo a physical state Y which causes behavior and internal functioning corresponding to a belief or judgment that there is a round thing in front of you (e.g., holding your hands in such-and-such way when trying to grab it, saying “this is round,” etc.). Conveniently, the psychophysical laws map Y onto an experience involving a phenomenal presentation of a round object, an experience that essentially provides you with justification to believe that there is a round object in front of you. The psychophysical laws map Y onto an experience whose essential normative role nicely harmonizes with the functional role of Y.

Again, it’s not puzzling that there is a physical state with this functional role. This presumably has an evolutionary explanation. What’s striking is that the actual psychophysical laws correlate it with an experience whose essential normative role harmonizes with this functional role. Holding fixed the physical and functional facts, the psychophysical laws could conceivably have mapped Y onto any number of alternative experiences—for example, a phenomenal presentation of a triangular object, or some random static. Then there would not have been a harmonious correspondence between our behavior/functioning and the reasons provided by our experiences.

Goff (2018) develops a related, partially overlapping puzzle, which he calls “the cognitive fine-tuning problem.” This is a puzzle specifically for proponents of cognitive phenomenalism—the view, roughly, that thoughts (e.g., occurrent beliefs and desires) are

constituted by non-sensory phenomenal states, where these phenomenal states are not grounded in purely functional states. In effect, the puzzle is to explain why cognitive phenomenal states exhibit a kind of normative harmony, which we might call *cognitive* harmony. Goff (2018: 113) writes, “The cognitive phenomenalist is obliged to give an explanation of why, of all the ways cognitive phenomenal states and sensory/functional states might have been matched, they tend to be matched in rationally appropriate ways.” One type of “rationally appropriate matching” involves practical rationality: we tend to behave in ways that are instrumentally rational in light of our (occurrent, phenomenal) beliefs and desires. Suppose your overall functional state involves a disposition to take a beer from the fridge. Given cognitive phenomenalism, you will also have some cognitive phenomenal states: perhaps among them a phenomenal belief that there is beer in the fridge and a phenomenal desire to drink beer. Although these cognitive phenomenal states are, given dualism, not grounded in your purely functional states, these superadded phenomenal states are exactly the states that would rationalize your behavior. Another type of rationally appropriate matching concerns theoretical rationality: we tend to have (occurrent phenomenal) beliefs that are rationally appropriate in light of our sensory experiences. For example, when you have a sensory experience of a table in front of you, this tends to be followed by a distinct phenomenal belief that there is a table in front of you. Of all the phenomenal beliefs that could have been correlated with your brain state by the psychophysical laws, you get a phenomenal belief that is rationally appropriate in light of your sensory experience.

Given cognitive phenomenalism, the fact that cognitive phenomenal states tend to be matched with sensory and functional states in such rationally appropriate ways calls out for explanation. Goff argues that, if cognitive phenomenalism is true, cognitive fine-tuning is difficult to explain naturalistically, but can be explained by theism or other non-naturalistic

views (e.g., “value-involving laws” (116-8)). We find this conditional plausible, but its antecedent—cognitive phenomenalism—is very controversial. Fortunately, we can afford to remain neutral on the truth of cognitive phenomenalism. If cognitive phenomenalism is true, then Goff’s examples of cognitive harmony are more grist for our mill. But if not, there are plenty of other examples of psychophysical harmony that don’t depend on it.

As Pautz and Goff emphasize, normative harmony cries out for explanation. We’ve seen that there does not appear to be any straightforward evolutionary explanation of normative harmony, but perhaps there are other naturalistic explanations. Here we will consider two naturalism-friendly responses to the alleged datum of psychophysical harmony. The first, which is not so much an explanation of the datum as a denial of it, is to adopt *normative error theory*. On this view, there is no correspondence between the normative role of an experience and its functional role because there are no normative facts at all. Pain is not bad, and does not give one reason to avoid and eliminate it; visual experiences do not justify beliefs; and so on. We find this view extremely implausible. It seems about as self-evident as anything in philosophy that excruciating pain is bad, or that experiences justify certain beliefs. If denying these claims is the price of upholding naturalistic atheism, the price is too high.

The second response, which we’ll call the *contingent normative roles explanation*, is more promising. According to this account, experiences have their normative roles contingently, in virtue of their contingently associated functional roles. If experiences have their *normative* roles in virtue of their *functional* roles, then arguably, it is no coincidence that the normative roles of experiences neatly align with their functional roles (cf. Pautz 2020: 9-10). Above we suggested that pain *essentially* is bad and provides one with reason to eliminate it. But a proponent of this response might suggest that the phenomenal state we call “pain” isn’t

essentially bad, but is only bad because we dislike it, or because of its association with avoidance behavior. This view may pair naturally with attitudinal theories of unpleasantness, according to which an experience is unpleasant in virtue of the attitude we take toward it, not in virtue of possessing a certain intrinsic phenomenal feature (Armstrong 1968, Parfit 1984, Heathwood 2007).

We can distinguish two versions of the contingent normative roles explanation. Focusing on pain for simplicity, the first says that pain is bad (when it is bad) in virtue of its phenomenal functional role, a role defined in terms of functional relations to other phenomenal states. For example, one might hold that pain is bad (when it is bad) because it gives rise to a state of “felt dislike,” an essentially phenomenal attitude (perhaps an instance of cognitive phenomenology).⁴ This version doesn’t solve our puzzle, but merely relocates it. Now the question is: why is the badness-conferring phenomenal state (e.g., felt dislike) linked to behavioral/functional properties (e.g., dispositions toward avoidance behavior) in harmonious ways?

The second version of the contingent normative roles explanation says that pain is bad in virtue of its non-phenomenal functional role, a role defined in non-phenomenal terms, ultimately in terms of causal relations (however indirect) to outward behavior and physical stimuli. (More precisely: a role defined in terms of causal relations to behavior, physical stimuli, and perhaps other internal states under “topic-neutral” descriptions, i.e., states that are characterized only in terms of their causal relations to behavior, physical stimuli, and each other.) Unlike the first version, this one doesn’t merely relocate the puzzle. However, this view is implausible on reflection. Imagine someone being tortured for an hour. The experience is non-instrumentally bad, an experience the subject has reason to eliminate. According to the view under

⁴ Kahane (2009) defends a view along these lines, and Aydede (2018) and Lin (2020) defend the closely related view that the pleasantness or unpleasantness of an experience depends on the phenomenal character of the attitude we take toward it.

consideration, what makes her experience bad is its functional connections to things “outside” her experience—ultimately, certain kinds of behavior and physical stimuli. To see why this is implausible, consider a disembodied Cartesian mind whose overall experience is exactly the same as that of the torture victim. The experience of the disembodied mind has no causal ties to any physical behavior or physical stimuli (because she has no body or physical sense organs). Still, it seems self-evident that her experience is bad, one she has reason to avoid. We submit that it is not even conceivable that there should be a perfect phenomenal duplicate of the torture victim whose experience is not intrinsically bad. (It may help to consider this point from the perspective of Cartesian doubt. When you’re in intense pain, you may be unsure whether you’re in a Cartesian skeptical scenario, with all your actual experiences but no physical body. But it’s clear that, even if you are in such a scenario, your pain is still bad.) For this reason, we think it is implausible that the normative role of experience is grounded in factors outside the phenomenal domain, such as functional ties to physical behavior and stimuli.

Similar points hold for the role of experience in providing epistemic reasons for belief. Much of the epistemic normative role of experience seems to depend only on factors within the phenomenal domain. For example, when we have a visual experience as of a sphere, the phenomenal character of the experience seems to essentially consist in ostensibly being presented with a spherical item, and it’s plausible that this necessarily provides us with defeasible reason to believe that there is a sphere before us. At a minimum, it seems to be a necessary consequence of our overall experience that we are justified in believing certain things about the distribution of sensible properties in our environments. For example, any lifelong phenomenal duplicate of you is very plausibly justified in believing that there are hand-shaped items at her side, even if she’s really a brain in a vat or a demon-deceived disembodied soul.

Thus, it is implausible to suppose that the epistemic normative role of experience is wholly grounded in its functional ties to things outside the phenomenal domain.

But suppose that one of the responses above (the contingent normative roles explanation or normative error theory) is successful. There is at least one other type of psychophysical harmony which is largely immune to these objections.

2.3 Semantic harmony

This other type is:

Semantic harmony: In many cases, the psychophysical laws pair phenomenal states with physical states in a way that generates a semantic correspondence between our judgments/reports and our phenomenal states.

We can illustrate the idea of semantic harmony by considering the “meta-problem of consciousness,” the focus of a number of recent papers. Recall Chalmers’ (1996) famous distinction between the “hard” and “easy” problems of consciousness. The easy problems involve explaining the behavior and functioning associated with consciousness (verbal reports, perceptual discrimination, learning, etc.). The hard problem involves explaining how and why physical activity in the brain gives rise to subjective experience. The latter is hard because facts about subjective experience seem to be “further facts” over and above any purely physical facts. Any description of your brain including only its purely physical operations seems compatible with the absence of consciousness, or with qualitatively different forms of consciousness.

The meta-problem is, very roughly, the problem of explaining why we think and say that there’s a hard problem of consciousness. The meta-problem counts as an “easy problem” because the explanandum is limited to behavior and functioning—e.g., the fact that we say things like “consciousness is something over and above information processing in the brain,” or “any purely

physical process is compatible with the absence of experience,” and make the corresponding judgments (in a functional/computational sense of “judgment”). Given causal completeness, there is a complete explanation in physical terms of the relevant problem reports and judgments which doesn’t reference the experiences that make these reports and judgments true. (Since we’re assuming dualism, we’re assuming these claims *are* true.) This seems very lucky. If there hadn’t been psychophysical laws correlating our physical states with distinct, non-physical states of consciousness, we would have made the same reports and judgments, but they would have been false. As Chalmers (2018: 48) remarks, “It is easy to get the sense that what really explains the intuitions is the structure of cognitive processes, and the fact that consciousness is connected to that structure is something of a fortunate and optional extra.”

The apparent luckiness here extends to a wide range of other judgments/reports we are inclined to make about the nature of our conscious states upon careful reflection. For example, when we have an experience as of a red and round thing, many of us are inclined to judge and report that our experience essentially involves an acquaintance relation with redness and roundness (Chalmers 2006, Tye 2008, Pautz 2010). When we have a reddish experience, an orangeish experience, and a greenish experience, we are disposed to judge that the first experience is more similar to the second than to the third. When we experience aesthetic delight, many of us are disposed to judge and report that this experience is intrinsically valuable. When we undergo a visual presentation of a cube, many of us are disposed to judge and report that we are in a state that essentially provides justification to believe that there is a cube before us.

In our view, these judgments are all very plausible, and many philosophers accept them. But if we grant the truth of these judgments, then—given dualism and causal completeness—we apparently have a very striking coincidence on our hands. Our underlying brain states dispose us

to make certain judgments and reports about the nature of our current phenomenal states. The psychophysical laws correlate these brain states with phenomenal states that make these very judgments and reports true. Yet there is a complete physical explanation of our phenomenal judgments/reports that makes no reference to our phenomenal states.

Semantic harmony is distinct from normative harmony, though the examples above may also involve normative harmony. The relevant reports and judgments may be *both* justified by the associated experiences *and* made true by them, but these are different things. A normative error theorist could agree that our judgments are *true* when we say that consciousness is non-physical, that we are acquainted with redness and roundness, or that reddish experiences are more similar to orangish experiences than to greenish ones, even though the error theorist will deny that anything justifies anything.

Granted, *some* of the semantic correspondence between our introspective judgments/reports and our experiences can probably be explained by the fact that our experiences help determine the meanings of our introspective reports and the contents of our introspective judgments. Consider a world physically just like ours, but in which our counterparts' experiences are spectrum-inverted relative to our own. When they look at a ripe tomato, they say things like, "I am having a phenomenally reddish experience," even though they are having (what we would call) a phenomenally greenish experience. Still, it's plausible that they say something true because, for them, "phenomenally reddish" refers to phenomenal greenishness. It's not as though we're lucky not to be spectrum inverted. But even though experience plausibly plays *some* role in determining the meanings of our introspective reports and the contents of our introspective judgments, this does not remove the apparent luckiness involved in the examples of semantic harmony above. Our statements and judgments about

phenomenal similarity (“this experience is more similar to this other experience than that experience”) would have been false under many possible psychophysical laws, such as laws that permute our reddish and orangeish experiences or laws that map all physical states onto the same buzzing-noise experience. Likewise, statements and judgments about the structure of experience, like “this experience involves an acquaintance relation with redness and roundness,” would presumably have been false if the psychophysical laws had given us an ear-ringing experience in place of a phenomenal presentation of a red and round object. And our statements and judgments about the value or disvalue of our experiences (“this experience is intrinsically good/bad”) would plausibly have been false in pleasure/pain inversion scenarios, and in scenarios where an evaluatively neutral state plays the pain or pleasure role.

Nor can we remove the sense of coincidence merely by saying that our phenomenal states are nomologically connected to our phenomenal reports/judgments in virtue of having a *common cause* in the brain (as they surely do). A common cause may explain why there is a *correlation* between our phenomenal states and our behavior/functioning. But it won’t by itself account for the *semantic correspondence* at issue. Consider two machines:

Text machine 1: Machine 1 outputs text and makes sounds. The text and sound outputs are correlated. For example, just before outputting “HIGH,” it always makes a crunching sound. Just before outputting “LOUD,” it emits a soft buzz. Just before outputting “RISING,” it emits a humming noise. And so on, for many other text/sound pairings.

Text machine 2: Machine 2 outputs text and makes sounds. The text and sound outputs are not only correlated, but also in semantic correspondence with one another. For example, just before outputting “HIGH,” it makes a high pitched noise. Just before

outputting “LOUD,” it makes a loud noise. Just before outputting “RISING,” it emits a rising tone. And so on, for many other sound/text pairings.

In Text Machine 1, we have a correlation between text outputs and sounds, but no other interesting correspondence between them beyond their regular co-occurrence. We can easily remove any sense of coincidence just by proposing that there is some common cause internal to the machine that is responsible for both the text output and the correlated sound. But things are different with Text Machine 2. We can't explain the semantic correspondence between text and sound, or remove the sense of coincidence, *merely* by saying that the text outputs and the sounds have a common cause. Something more is needed. In the same way, the mere existence of a common cause for our phenomenal states and our phenomenal judgments/reports does not explain semantic harmony.⁵

2.4 From harmony to theism

So: the psychophysical laws ensure that phenomenal states pattern with each other, and with other physical states, in harmonious ways, ways that seem extremely lucky. The overwhelming majority of sets of conceivable psychophysical laws would have produced disharmony, as would the absence of consciousness altogether. It also seems that the *simplest* (and hence most intrinsically probable) psychophysical correlation patterns would be

⁵ There may be a parallel with “third-factor responses” to evolutionary attempts to debunk moral knowledge (e.g., Enoch 2010; Wielenberg 2010). These attempts claim that the availability of a complete evolutionary explanation of our moral faculties which doesn't reference the reliability of these faculties means that we don't have moral knowledge. Third-factor explanations attempt to find some factor which is responsible for both the fact that actions have the moral status they do and the fact that we think that they have this moral status, and which ensures that there's at least some correspondence between our beliefs and the moral facts. Suppose our belief that X is morally good doesn't explain why it's morally good, and that X's moral goodness doesn't explain why we believe it's morally good. But also suppose that (i) our faculties were selected to make us think that survival-promoting things are morally good, and that (ii) actually, survival-promoting things usually are good. Assuming our faculties couldn't have easily been selected without regard for survival, and assuming that survival couldn't have easily failed to be good, it may be unsurprising that there's some correspondence between our beliefs and the moral facts. Critics (e.g., Korman and Locke forthcoming) have pursued various responses. We're suggesting that the common cause explanation here fails to even get as far as the third-factor debunking argument response: it doesn't even significantly raise the probability that semantic correspondence holds. We leave open whether criticisms of third-factor responses might pose further problems for common cause solutions to semantic correspondence.

disharmonious. For example, the psychophysical laws would be simpler if they mapped all physical states onto the same experience of buzzing noise. Relatedly, it's striking that the psychophysical laws, unlike other fundamental laws, appear to operate on relatively macroscopic physical states, such as neural firing patterns or high-level information structures in the brain. As J.J.C. Smart (1959: 143) famously remarked, it is this feature of the dualist's ultimate laws that give them an odd "smell," unlike anything else known to science (cf. Pautz forthcoming). We might have expected the psychophysical laws to instead be directly sensitive to microphysical phenomena, with our conscious experiences somehow mirroring the microphysical structure of our brains. But that would result in a chaotic and disharmonious mess.

On ordinary naturalistic atheism, it's very hard to see what could explain psychophysical harmony. On theism, it's easy. Psychophysical harmony is valuable, inasmuch as it allows there to be beings who not only have phenomenal states, but have phenomenal states which can play normatively appropriate roles, who possess meaningful agency, who can respond rationally to sensory evidence, who can behave rationally on the basis of their desires, and who can have reasonably reliable intuitions about their phenomenal states. Given the value of psychophysical harmony, it's not terribly unlikely that God would create a world whose laws are fine-tuned for psychophysical harmony. Pautz (2020: 6) even rhetorically asks, "What—short of an intelligent designer—might explain why the psychological laws are actually 'fine-tuned' to result in normative harmony?" Several other authors (e.g., Goff 2018: 116; Latham 2000: 76; Mørch 2017: 298) also raise the possibility of a theistic explanation of psychophysical harmony (though they typically quickly dismiss such an explanation as otherwise unacceptable, a fact we'll return to later).

Let's formulate the argument more precisely in Bayesian terms. This involves comparing the likelihood of harmony on rival hypotheses. In what follows, assume for simplicity that the rejection of theism entails ordinary naturalistic atheism. This is obviously false, especially given our earlier remarks about theism-adjacent hypotheses. But if the argument shows that the real choice is between theism and theism-adjacent hypotheses, that itself is hugely significant.

Given that the vast majority of conceivable psychophysical mappings would be disharmonious, and especially if (as we suggested above) the individually most intrinsically probable mappings are disharmonious, the epistemic probability of harmony given (ordinary naturalistic) atheism seems *extremely* low. On the other hand, for reasons given above, it's far more likely that there would be psychophysical harmony conditional on theism. That is:

Likelihood Comparison: $P(\text{harmony}|\text{theism}) \gg P(\text{harmony}|\text{atheism})$

Unless the prior probability of theism is fantastically low, the posterior probability of theism will be reasonably high when we conditionalize on psychophysical harmony.

This reasoning has the same probabilistic structure as other cases where we respond to a fact that “cries out for explanation” by revising our assumptions about what brought it about. A monkey we assumed to be typing at random types out “methinks it is like a weasel.” A student we assumed to be honest turns in a paper that matches an SEP article verbatim. A coin we assumed to be fair lands heads 15 times in a row. Laws of nature we assumed to be undesigned turn out to be strikingly harmonious. In each case, we observe a fact such that (i) the fact is very improbable given our initial assumption about the circumstances that brought about the fact in question (e.g., that the student is acting honestly, that the basic laws of the universe are undesigned and purposeless), and (ii) there is an alternative to our initial assumption that is not fantastically unlikely that makes the observed fact much more likely (e.g., that the student is

cheating, that the laws are the product of design). We then respond to the evidence by significantly raising our confidence in this alternative hypothesis (cf. White 2005).

Three elementary points about the Bayesian framework are worth noting. First, the improbability of harmony on atheism is not *itself* especially significant. What matters is that it's extremely improbable *compared to* the probability of harmony on theism. The specific arrangement of stars in the night sky is very unlikely on atheism. But it's equally unlikely on theism, and so not evidence for it. On the other hand, stars arranged to spell out "I am the Alpha and the Omega" *would* be evidence for theism. This would be no less likely on atheism than any comparably specific arrangement, including the actual one. What matters is that this arrangement is *much more* likely on theism than on atheism.

More precisely, what determines the evidential impact of a piece of evidence e is the likelihood ratio: $P(e|\text{theism})/P(e|\text{atheism})$. Whatever this ratio is, multiplying it by the ratio of the prior probability of theism to the prior probability of atheism gives us the new ratio of the probability of theism to the probability of atheism. To illustrate with some artificial numbers: suppose that harmony is 50 times more likely on theism than on atheism. (We think that's a *very* conservative estimate.) And suppose that *before* considering psychophysical harmony, theism is 1/10 as probable as atheism (i.e., atheism is 10 times more probable than theism). Then *after* conditioning on psychophysical harmony, one should think that theism is $1/10 * 50 = 5$ times as likely as atheism.

A second, closely related point is that that $P(\text{harmony}|\text{theism})$ need not be high in any "absolute" sense, provided it's significantly higher than $P(\text{harmony}|\text{atheism})$. Again, the Alpha/Omega case is instructive. Given theism, the Alpha/Omega Pattern does not seem very likely in any absolute sense. It would surely be unreasonable to assign a probability higher than

1/10,000 to that specific arrangement, conditional on theism. (If that's not obvious, make the message longer and more specific.) 1/10,000 sounds like a small number, but even if the correct value is much lower than 1/10,000, it will be many orders of magnitude higher than $P(\text{alpha-arrangement}|\text{atheism})$. This is why it is such powerful evidence. We've provided reasons (having to do with the value of psychophysical harmony) for thinking that psychophysical harmony isn't terribly surprising given theism. Perhaps they don't show that $P(\text{harmony}|\text{theism})$ is very high. We accept the familiar point that divine psychology is hard. We should be modest in our judgments about what God would do. The key point is that we don't need to assume that $P(\text{harmony}|\text{theism})$ is high in any absolute sense, just that it is much higher than $P(\text{harmony}|\text{atheism})$.

Finally, Likelihood Comparison is only interesting if the prior probability of theism is not so fantastically low that its posterior probability is negligible even after the evidential boost it receives from psychophysical harmony. We noted above that several authors acknowledge but quickly dismiss the possibility of a theistic explanation of psychophysical harmony. Sometimes the dismissal comes even when the author acknowledges psychophysical harmony as a serious problem and lacks any other explanation for it. For instance, Adam Pautz (2020: 13) writes that "I don't have a good response to the normative harmony problem... I consider it to be a deeply troubling but largely overlooked element of the mind-body problem." However, he dismisses a theistic explanation:

[Dualists] could say that the fact that the psychophysical laws are harmonious, when added to the other wonderful facts about the universe (puppies, iPhones, etc.), provides enough evidence of a harmony-loving God that we should believe in such a God. But it's far from obvious that such a conclusion would be warranted (especially if the rational "prior probability" of such a God is extremely low to begin with). (Why not believe in a cheese-loving God whose main goal was to create cheese, given the plenitude of cheese on earth? Or a suffering-loving

God, given all the suffering?)⁶ And, anyway, this theistic view is complex and objectionable in other ways. (ms.: 40-1)

Discussants often suppose that atheism is a precondition on a successful explanation of harmony. If the prior probability of theism is fantastically low, this is justified—but not otherwise.

We think it is unreasonable to assign theism such a fantastically low prior. The prior probability of theism is the result of two factors: (i) its intrinsic probability, i.e., its probability conditional on no evidence, and (ii) its fit with our background knowledge (cf. Swinburne 2004, chs. 1-5). The intrinsic probability of theism doesn't seem prohibitively low. It claims that the foundation of reality consists of a mind which possesses a few reasonably natural features (power, goodness, knowledge) to a reasonably natural degree (i.e., a maximal degree) (cf. Swinburne 2004, ch. 5). (Some authors even claim that God has just a *single* fundamental property held to a *maximally* natural degree.)⁷ Simplicity and naturalness or non-arbitrariness are plausibly key determinants of intrinsic probability.⁸

Theism thus differs from other design hypotheses that *should* be assigned extremely low intrinsic probabilities in virtue of their extreme arbitrariness and specificity. For example, let S-theism be the hypothesis that there is a God whose favorite arrangement is S, where S is the actual arrangement of stars we observe in the night sky. The likelihood ratio $P(S|S\text{-theism})/P(S|\sim S\text{-theism})$ is staggeringly large. But it would be unreasonable to accept S-theism even after taking account of the S-arrangement of stars. This can only be because the prior of S-theism is low to a correspondingly staggering degree. Part of the explanation for S-theism's low prior probability is the existence of a staggering number of alternative design hypotheses (e.g., one for each separate arrangement of stars) each of which seems no less

⁶ The existence of evil is also cited in this context by Mørch (2017: 311 fn. 1).

⁷ E.g., for Rasmussen (see Rasmussen and Leon 2019) this is perfection or maximal value, and for Swinburne 2009 it is "pure, limitless, intentional power."

⁸ Though for an approach which explicitly eschews appeals to simplicity while nonetheless concluding that theism has a reasonably high intrinsic probability, see Poston (forthcoming).

initially probable than S-theism. Since these hypotheses are mutually exclusive and extremely numerous, any single such hypothesis receives only a sliver of probability.

In some sense, traditional theism is also very specific. It excludes the hypotheses that God cares only about producing cheese, that God knows everything except the 837,843,426th digit of pi, etc. But scientific hypotheses are often specific in the same way: e.g., the hypothesis that X is a universal law of nature excludes the hypotheses that X holds everywhere except a one-inch patch on the far side of EBLM JO555-57, that it holds only until tomorrow at one second past noon, or 1.1 seconds, or 1.11 seconds... Since these alternatives fit equally well with our observations, our accepting the hypothesis that X is a universal law rather than the disjunction of the infinitely many alternatives it excludes must be based on its greater theoretical virtue—unlike S-theism, its competitors are *not* individually as probable as it—and this will have something to do with its comparative simplicity and naturalness. However, it's plausible that assigning, say, omniscience to God is similarly simpler and more natural than assigning knowledge of everything except a random digit of pi, or some other arbitrarily limited amount of knowledge (cf. Miller 2016).

Now consider the cheese-loving God. It seems reasonably simple and natural to suppose that God would be motivated to act in accordance with what God takes God's reasons to be (this is even entailed by relevant forms of motivational internalism, which presumably have some non-negligible epistemic probability of being true). If God is omniscient, God will know what God's reasons actually are. And it seems plausible that everyone, including God, has reason to produce things of value. From these suppositions, it follows that God will be motivated to produce things of value, as we've assumed (cf. Swinburne 2004, ch. 5). There is no parallel argument for supposing God would have a completely arbitrary motivation like only caring about

producing cheese. Meanwhile, Pautz's suggestion that we might as well posit a "suffering loving-God, given all the suffering" parallels Stephen Law's (2010) "evil-god challenge." The challenge has attracted a number of responses (e.g., Forrest 2012; Ward 2015; Weaver 2015; Hendricks 2018; Rasmussen 2019, esp. 254-257; Page and Baker-Hytch 2020; Milburn forthcoming; Miller forthcoming; Poston forthcoming, sec. 4.1), some of which cover similar points as those above.

In the present context, perhaps our background knowledge should be taken to include the data appealed to by various other arguments for and against theism, including the existence of suffering. We need not take a detailed stance on the strength of these arguments. Even if our background knowledge favors atheism, it's enough for us if it doesn't render the prior probability of theism *prohibitively* low. We think this is reasonable, particularly given that, to render the posterior probability of theism negligible, the evidence for atheism needs to significantly outweigh not only the "standard" evidence for theism but also the significant evidence provided by psychophysical harmony. Suppose evil outweighs the standard theistic evidence. We think it's plausible that psychophysical harmony is orders of magnitude more likely on theism than atheism. In that case, even a very small probability that some theodicy or other works might be enough for harmony to more than outweigh evil.

Moreover, even if one thinks the prior probability of *theism* is prohibitively low, one may not think the prior probability of all theism-adjacent hypotheses is prohibitively low. For example, Draper's (2017) "aesthetic deism" is formulated with the specific aim of better accounting for evil, whereas process theism maintains God's moral perfection but constrains God's power (Griffin 2004). Someone who thinks evil undermines theism may still find our

argument interesting if framed, not as an argument for theism, but rather as an argument for the disjunction of theism and all relevant theism-adjacent hypotheses.

3. Causal completeness and dualism are inessential to the argument

3.1 We don't need to assume causal completeness

Above, we made two controversial metaphysical assumptions: dualism and causal completeness. While these assumptions were convenient for an initial presentation of the argument, we argue in this section that they are ultimately inessential. The argument from psychophysical harmony for theism retains its force even if we drop these assumptions.

Let us start with causal completeness. Interactionist dualists reject causal completeness, claiming that phenomenal states have a non-redundant influence on physical processing in the brain. But interactionism by itself does not explain psychophysical harmony or render it unsurprising. Recall Text Machine 2, the machine whose text output semantically corresponds with its sound output. One might hypothesize that the sound outputs exert some causal influence on text outputs. But this hypothesis, by itself, would not lead us to expect a semantic correspondence between text and sound. After all, there are countless ways in which text output might causally depend on sound, most of which do not push the text in the direction of semantic correspondence. Higher pitched sounds might have biased the machine toward outputting text near the end of the alphabet; louder sounds might have biased the machine toward longer text outputs; the timing of the text outputs might have tracked the rhythm of the sounds. These dependence patterns would have made the text-sound relationship somewhat *different* from what it would have been in the absence of sound-to-text influence, but (in general) not any closer to semantic correspondence. The same goes for countless other natural rules that might have characterized the causal dependence of text on sound. If the sounds *do* bias the text output in the

direction of semantic correspondence, it is because the rules governing the dependence of text on sound take a very special and distinctive form.

Similarly, there are many ways in which physical states might causally depend on phenomenal states. Absent design, there is no reason to expect that the actual pattern would push physical processing in the direction of psychophysical harmony. The laws governing phenomenal-to-physical causation could just as well have pushed physical processing away from harmony or in a direction orthogonal to it. The psychophysical laws would have to take a very special and distinctive form in order to achieve harmony, and it would be surprising for the laws to take this special form if they are undesigned, while this would be relatively unsurprising on theism. The point can be illustrated with some examples. If interactionism is true, the actual interactionist laws assign something like the following nomological role to pain: a certain brain state X causes pain, and pain in turn causes avoidance behavior. Here we have a case of hedonic harmony. But if the interactionist laws had been different, this nomological role might have been occupied by pleasure (with brain state X causing pleasure, and pleasure causing avoidance behavior) or by some evaluatively neutral state. In that case, we'd have hedonic *disharmony*.⁹ Likewise, if interactionism is true, the actual interactionist laws assign a certain nomological role to sphere-ish visual experiences, resulting in an instance of epistemic harmony. But if the interactionist laws had assigned this nomological role to any number of different phenomenal states (e.g., a phenomenal presentation of a triangle, or a random-static experience), there would have been epistemic *disharmony*. The fact that the interactionist laws assign nomic roles to experiences in ways that yield psychophysical harmony doesn't come for free on interactionism. It requires explanation.

⁹ Robinson (2007) and Corabi (2014) make a similar point in response to William James's evolutionary argument against epiphenomenalist dualism.

The most promising interactionist proposal we know of that aims to account for psychophysical harmony is Bradford Saad's (2019) proposed solution to Chalmers' meta-problem. Saad proposes that a fundamental teleological law of nature operates on normative features of experiences to ensure that physical processing is biased toward physical states that are rationalized by our experiences. More specifically, Saad proposes that it is a fundamental law of nature that an experience causes whatever physical state is the most normatively appropriate response to that experience (subject to certain constraints).¹⁰ We have no serious objections to Saad's proposal, at least if it is taken (as he intends) as a programmatic sketch of what one of the ultimate psychophysical laws might be like. However, if Saad's proposal is correct—if the universe is governed by fundamental teleological laws that explicitly involve normative notions and direct the world toward normatively favored outcomes—this itself seems to be evidence for theism. Moreover, even if one combines Saad's proposal with atheism, treating these teleological laws as explanatorily basic, the resulting view stands in conflict with our main target: naturalistic atheism. According to the latter, the universe is not fundamentally ordered toward the realization of any kind of normatively favored outcome. Saad's view is akin to Nagel's anti-naturalistic view that there are basic teleological laws that order the universe toward the realization of certain values. (Saad (2019: 208n) notes the close relationship between Nagel's view and his own.) It is no part of our goal in this paper to argue that theism is preferable to theism-adjacent views of this kind.¹¹

¹⁰ Goff (2018: 116-8) sympathetically discusses a similar view as a potential solution to his cognitive fine-tuning puzzle.

¹¹ It's worth noting that there is a non-interactionist variation on Saad's proposal that might be embraced by dualists who accept causal completeness. Instead of a "top-down" law, there might just be a "bottom-up" psychophysical law governing physical-to-phenomenal causation which says that physical states generate the phenomenal state whose essential normative role best matches the functional role of the brain state (perhaps subject to certain constraints). Thus, we think that Pautz (2020) may be wrong to suggest that Saad's proposal shows that interactionist dualism is better positioned to solve the puzzle of psychophysical luck than non-interactionist dualism. If one is willing to go in for normativity-laden teleological laws, the bottom-up route may be just as effective at achieving psychophysical harmony as Saad's top-down strategy.

The argument above assumes that there are many conceivable ways (that is, many *a priori* epistemically possible ways) for phenomenal states to influence physical processing. However, Hedda Hassel Mørch (2017, 2020) argues that there are necessary, synthetic *a priori* truths about the causal powers of phenomenal states, such that it's not even conceivable for phenomenal states to have causal powers different from the powers they actually have. While Mørch allows for the conceivability of scenarios in which pain (for example) has *no* causal powers, she maintains that it is *a priori* that, if pain has any causal power, it has the power of disposing one to try to eliminate it. Harold Langsam (2011) makes similar claims about the causal powers of sensory experiences to produce thoughts and beliefs about sensible properties. Notably, Mørch's and Langsam's examples of *a priori* causal powers all involve the powers of phenomenal states to produce certain *mental* effects rather than physical effects (e.g., *trying* to eliminate pain rather than behaving in ways that eliminate pain; *thinking* about redness rather than saying "red"). We grant that it is at least difficult to conceive of phenomenal states having radically different causal powers with respect to some of their purely mental effects. However, we think that scenarios in which phenomenal states have radically different causal powers with respect to their *physical* effects are plainly conceivable. For example, there is no *a priori* incoherence in the claim that an experience of pain directly causes a candle to be extinguished on the other side of the world, or in the claim that an experience of pain causes particles to swerve in one's brain in ways that aren't conducive to pain avoidance.

3.2 We don't need to assume dualism

So far we've assumed that the phenomenal truths could have varied independently of the physical/functional truths. Physicalist views of consciousness deny this assumption. One might therefore think that the argument fails unless we assume that physicalism is false from the outset.

Even if this were correct, the argument would be significant, since many philosophers reject physicalism for reasons independent of their views on theism. However, we will argue that the success of the argument does not depend on a prior rejection of physicalism. The datum of psychophysical harmony provides strong evidence for theism even if one assigns a substantial prior probability to physicalist views of consciousness.

Everyone, even physicalists, should accept the datum of psychophysical harmony. (With one qualification: physicalists will reject one of the alleged cases of semantic harmony, namely, the alleged semantic correspondence between the non-physicality of consciousness and our judgments/reports about the non-physicality of consciousness. If physicalism is true, there is no semantic correspondence here, just an illusion of non-physicality. But physicalists can accept our other examples of semantic harmony, and should also accept the existence of normative harmony, which will be our focus here.) Since everyone should accept the datum of psychophysical harmony, the only way physicalism could be relevant to our argument is if physicalism has some bearing on the likelihoods involved in the key Likelihood Comparison:

Likelihood Comparison: $P(\text{harmony}|\text{theism}) \gg P(\text{harmony}|\text{atheism})$.

Although physicalism raises subtle and tricky issues, we shall argue that these likelihoods are probably not significantly affected if we stop treating dualism as part of our background knowledge, and allow for some prior probability that physicalism is true. However, the success of the argument probably *does* require that we not assign physicalism a prior probability of 1 (or too close to 1). This is because our conclusion that the psychophysical correlations are the product of design may entail the falsity of physicalism. But even if the falsity of physicalism falls out by the *end* of the argument, we needn't assume that physicalism is false or very improbable at the *beginning* of the argument.

We will assume, with most physicalists nowadays, that whether or not there is an *ontological* gap between physical/functional truths and phenomenal truths, there is at least an *epistemic* gap. In other words, scenarios in which phenomenal truths vary independently of the physical/functional truths are *a priori* epistemic possibilities, even if these don't correspond to any metaphysically possible scenario.¹² For example, we assume there are *a priori* epistemic possibilities in which pain is correlated with different (and normatively mismatched) functional properties. If it is epistemically possible that pain is identical to, or grounded in, some physical/functional property (as physicalists suppose), then presumably this epistemic possibility subsumes a range of more specific possibilities corresponding to the range of epistemically possible psychophysical correlations. In other words, there will be a range of distinct physical/functional properties that might (*a priori*) turn out to be identical to, or to ground, the phenomenal property of being in pain.

We can distinguish *identity* physicalism from *ground* physicalism. The former holds that phenomenal properties are identical to physical/functional properties. The latter holds that phenomenal properties are not identical to any physical/functional properties, but are grounded in physical/functional properties via *a posteriori* principles of psychophysical grounding.¹³ Our main focus in this section will be identity physicalism, but it will be worthwhile to briefly comment on the relevance of ground physicalism to our argument. First, note that ground physicalism seems not to have an effect on the right-hand side of our Likelihood Comparison (i.e., $P(\text{harmony}|\text{atheism})$). Ground physicalism differs from dualism in *metaphysical* respects: it replaces metaphysically contingent causal laws with metaphysically necessary laws of

¹² Thus, we assume the falsity of so-called "*a priori*" (or "type-A") physicalism, the view that physicalism is true and there is not even an epistemic gap between the physical truths and the phenomenal truths.

¹³ See Schaffer (forthcoming) for a recent defense of ground physicalism, and Pautz (forthcoming) for a useful discussion of the relative merits of identity physicalism and ground physicalism.

grounding. But in *epistemic* respects, the ground physicalist's *a posteriori* grounding laws seem not to be importantly different from the dualist's *a posteriori* causal laws. For just about any conceivable set of dualistic psychophysical laws, there is a corresponding conceivable set of grounding laws.¹⁴ As with the dualist's causal laws, there was no *a priori* guarantee that the psychophysical grounding laws would turn out to be harmony-inducing. And we see no reason why the epistemic probability of harmony-inducing grounding laws on atheistic ground physicalism should be any higher than the probability of harmony-inducing causal laws on atheistic dualism. Hence, allowing some prior probability for ground physicalism would seem not to affect the value of $P(\text{harmony}|\text{atheism})$. On the other hand, ground physicalism may have a modest effect on the other side of the Likelihood Comparison (i.e., $P(\text{harmony}|\text{theism})$). Below we will see that the same is true for identity physicalism, but we will argue that this does not undermine the argument. Since the discussion below applies equally to ground physicalism, we will not consider the issue separately here.

Let us then set aside ground physicalism and turn to identity physicalism. For simplicity, our discussion will focus on *functionalist* forms of identity physicalism, those that identify phenomenal properties with functional properties. The reason for focusing on functionalism is that it may seem to offer an especially straightforward solution to the puzzle of psychophysical harmony without invoking God. Consider normative harmony as it applies to pain. Our explanandum is, roughly, that pain is correlated with a certain aversive functional property—call it F—that normatively harmonizes with pain. If being in pain *is* F, this would seem to explain why they are correlated. One might object that this explanation is unsatisfying because the fact that pain is identical to F, rather than to some other functional property with which pain doesn't

¹⁴ Perhaps indeterministic dualist laws are an exception, since arguably there cannot be indeterministic grounding laws.

harmonize, would itself cry out for explanation. But the functionalist can respond, somewhat plausibly, that identity facts don't admit of (and therefore don't cry out for) explanation.

We do not deny that the identity of pain and F would explain harmony. Nor will we challenge the assumption that identity facts never cry out for explanation.¹⁵ But does this have any bearing on the likelihoods involved in our Likelihood Comparison? It is tempting to think that, if we treat functionalism as a serious epistemic possibility, then $P(\text{harmony}|\text{atheism})$ will be fairly high. The functionalist says that being in pain = F. Since harmony consists in the correlation between pain and F, this identification guarantees harmony. It therefore seems that $P(\text{harmony}|\text{atheism} \ \& \ \text{functionalism})$ should be equal to 1. In that case, $P(\text{harmony}|\text{atheism})$ will be reasonably high, provided that functionalism is at least reasonably probable given atheism. But if $P(\text{harmony}|\text{atheism})$ is high, our argument fails.

To see where the reasoning above goes wrong, consider a parody argument for the conclusion that $P(\text{harmony}|\text{atheism} \ \& \ \text{dualism}) = 1$.

The dualist holds that pain is nomologically linked to F. Since harmony consists in the correlation between pain and F, this nomological link guarantees harmony. So,

$P(\text{harmony}|\text{atheism} \ \& \ \text{dualism}) = 1$.

The problem with this argument is that it confuses the general thesis of dualism with a highly specific version of dualism. Dualism (about pain) does not imply that pain is nomologically linked to F in particular. Rather, dualism says (very roughly) that pain is nomologically linked to *some* physical/functional property. This general thesis subsumes a range of specific possibilities (“pain is nomologically linked to F1,” “pain is nomologically linked to F2,” and so on). At best, dualism implies that pain is nomologically linked to F only when taken in conjunction with *a*

¹⁵ Whether this assumption is true is a difficult question. See Chalmers (2020: 269) for the opposing view that *a posteriori* identities can call out for explanation, and that “harmonious identities” in particular are “potentially in need of explanation.”

posteriori facts about psycho-functional correlations. Now, if we treated these facts as part of background knowledge, then $P(\text{harmony}|\text{atheism} \ \& \ \text{dualism})$ would indeed be equal to 1. But it is crucial to the original argument that information about psycho-functional correlations is *not* treated as part of our background knowledge. If it were, then the *prior* probability of harmony would be 1, since the datum of harmony is itself a fact about psycho-functional correlations.

For the same reason, it would be a mistake to conclude on the basis of the reasoning above that $P(\text{harmony}|\text{atheism} \ \& \ \text{functionalism}) = 1$. Functionalism (about pain) is not the thesis that being in pain is identical to F in particular. It is the thesis that being in pain is identical to some functional property. As with dualism, functionalism subdivides into a range of more specific epistemic possibilities: pain = F1, pain = F2, and so forth. (These are a priori *epistemic* possibilities even if at most one is a *metaphysical* possibility.) At best, functionalism implies that being in pain = F only when taken in conjunction with *a posteriori* facts about psycho-functional correlations. Now, if we treated this information as part of background knowledge, then $P(\text{harmony}|\text{atheism} \ \& \ \text{functionalism})$ would be equal to 1. But, again, it is crucial to the original argument that information about psycho-functional correlations is *not* treated as part of our background knowledge. (Note that nothing substantive hinges on the verbal point that “functionalism” refers to the weaker claim that pain is *some* functional property rather than the stronger claim that pain = F. We can, if we like, use “functionalism” to refer to the stronger proposition. In that case, $P(\text{harmony}|\text{atheism} \ \& \ \text{functionalism}) = 1$. But then we can’t infer that $P(\text{harmony}|\text{atheism})$ is fairly high, because there’s no reason to think that this *highly specific* functionalist view should cover a large fraction of the atheist region of our probability space. In the same way, using “dualism” to express the specific thesis that pain is nomologically linked to

F in particular won't help us argue that $P(\text{harmony}|\text{atheism})$ is high, since there is no reason to think that this highly specific dualist view is very probable on atheism.)

It seems to us that the prior epistemic probability of harmony given [atheism & functionalism] should be more-or-less the same as the prior epistemic probability of harmony given [atheism & dualism]. In §2.4, we argued, in effect, that $P(\text{harmony}|\text{atheism \& dualism})$ is very low. The motivation for this, very roughly, was that the probability function $P(\cdot|\text{atheism \& dualism})$ distributes probabilities over a large set of conceivable correlation patterns, each corresponding to a hypothesis about what the psychophysical laws will look like, and most of these (including the simplest) are disharmonious. As we argued in §2.4, given atheism, harmonious patterns shouldn't receive a much greater probability than other comparably specific and simple disharmonious patterns, so $P(\text{harmony}|\text{atheism \& dualism})$ should be very low. Similarly, the probability function $P(\cdot|\text{atheism \& functionalism})$ distributes probabilities over more-or-less the same set of conceivable correlation patterns.¹⁶ In this case, each conceivable pattern corresponds to a hypothesis about how the *a posteriori* psycho-functional identities will turn out. But it's not clear why this should affect how we distribute epistemic probabilities across them. There's no obvious reason why harmonious correlation patterns should be assigned a much greater probability than other comparably specific and simple disharmonious patterns. In other words, there is no obvious reason why the *a priori* epistemic probability of a harmony-implying psycho-functional identity hypothesis (e.g., pain = aversive functional property F) should be much greater than that of a comparably specific disharmony-implying psycho-functional identity hypothesis (e.g., pain = non-aversive functional property F*).

¹⁶ A bit more precisely, any conceivable dualistic correlation pattern that yields a one-one mapping between phenomenal properties and some collection of functional properties will correspond to a conceivable functionalist scenario.

If the *a priori* epistemic probability of harmonious psychophysical identity hypotheses is no higher than that of comparably specific disharmonious identity hypotheses, then $P(\text{harmony}|\text{atheism} \ \& \ \text{functionalism})$ should be very low, more-or-less as low as $P(\text{harmony}|\text{atheism} \ \& \ \text{dualism})$. Let's assume for simplicity that dualism is the only relevant alternative to functionalism. (We consider other views below.) If these two likelihoods are (roughly) the same, it follows that allowing some prior probability for functionalism makes (roughly) no difference to the right side of the likelihood comparison. Whatever value one assigns to $P(\text{harmony}|\text{atheism} \ \& \ \text{dualism})$, roughly the same value should be assigned to $P(\text{harmony}|\text{atheism})$ regardless of one's prior attitudes toward functionalism.

If functionalism has no effect on $P(\text{harmony}|\text{atheism})$, then it can only be relevant to our Likelihood Comparison if it has an effect on the other side: $P(\text{harmony}|\text{theism})$. Here, functionalism probably has a modest effect. Functionalism seems to imply that God cannot control the psycho-functional correlation patterns (because God has no control over necessary truths, like psycho-functional identities). So, psychophysical harmony is presumably just as unlikely on [theism & functionalism] as on atheism. Even so, this won't affect the argument in any decisive way unless $P(\text{functionalism}|\text{theism})$ is very close to 1—that is, unless one is extremely confident in functionalism given theism. Even if (prior to considering psychophysical harmony) one is *pretty* confident in functionalism given theism, conditioning on psychophysical harmony should make one confident in theism *and* dualism.

Here's an analogy:

Boulder Island: You're about to visit the island of boulders. You don't know whether anyone has been there before. You're fairly (but not extremely) confident that boulders are not humanly moveable, so even if there were previous visitors, they couldn't have

controlled the arrangement of boulders on the island. (Boulders are very heavy, after all.)

When you arrive at the island, you observe the boulders arranged in a nice circular pattern, like Stonehenge.

Intuitively, you should now be fairly confident in the following conjunction: *there were previous visitors, and they could control the arrangement of the boulders*. The probability of the Stonehenge-like arrangement given this conjunction is so much higher than the probability of the Stonehenge-like arrangement given the falsity of this conjunction that, after updating, one should be very confident in the conjunction unless one had extremely low confidence in it to begin with. In just the same way, even if you have a high prior probability that functionalism is true (and hence that God can't control the psychophysical correlation patterns, even given theism), *after* taking account of psychophysical harmony, you should be much more confident that (i) God exists, and (ii) the psychophysical laws are subject to his control (so functionalism is false).

For the reasons just given, it's plausible that psychophysical harmony is actually evidence *against* the functionalist thesis that phenomenal properties are identical to functional properties. This, of course, is not to deny that it is evidence for a certain specific version of functionalism, such as one that identifies being in pain with the aversive functional property F. But harmony can be evidence against functionalism (and for dualistic theism) even if it is evidence for a certain specific version of functionalism. (By analogy, the Scranton Strangler's presence at the murder scene lowers the probability that the victim was poisoned, but raises the probability that he was poisoned by the Scranton Strangler.) There are some subtleties here, though. If we don't take it as part of our background knowledge that pain is correlated with some functional state, then this fact must be regarded as evidence for functionalism (but not strong evidence for functionalism over and against the design hypothesis, which also predicts regular correlations). Even so, the

more specific discovery that pain correlates with F in particular might be evidence against functionalism. And, indeed, we think that is the case here.

Another analogy will be helpful. Suppose we are ancient astronomers wondering whether Hesperus is identical to any of the many “other” (conceptually distinct) astronomical bodies we know about: $A_1, A_2, A_3 \dots A_n$. The Identity Hypothesis says that Hesperus is identical to one of the A s, while the Distinctness Hypothesis denies this. The Identity Hypothesis subdivides into various specific hypotheses: “Hesperus = A_1 ,” “Hesperus = A_2 ,” and so on. Suppose we discover that Hesperus “matches” A_{517} in certain respects, e.g., they have the exact same size, mass, and spectral composition. This discovery is, of course, evidence for the specific identity hypothesis “Hesperus = A_{517} .” Is it also evidence for the general hypothesis? It depends. If it is not part of background knowledge that there would be such a match between Hesperus and one of the A s, and if there is nothing special about A_{517} , then this match would certainly strongly confirm the general Identity Hypothesis. But one can imagine special circumstances—circumstances in which A_{517} is special among the A s in a certain way—where this discovery would *disconfirm* the Identity Hypothesis, and these circumstances are importantly analogous to those that hold in the case of psychophysical harmony. Imagine that “Hesperus matches A_{517} ” is special among the various *a priori* possible matching scenarios (i.e., “Hesperus matches A_1 ,” “Hesperus matches A_2 ,” and so forth) in the following way: There is a certain not-very-implausible theory that entails the Distinctness Hypothesis and predicts a match *specifically with* A_{517} , but the same cannot be said for any of the other A s besides A_{517} (no respectable theory entails Distinctness and predicts a match with A_{516} , for example). Suppose further that among the various specific identity hypotheses, “Hesperus = A_{517} ” does not stand out from the crowd as being antecedently more likely than the others.

More concretely, we might suppose that a central teaching of the local religion, which we take somewhat seriously, is that the Mother Goddess created the Morning Star (A_{517}) and the Father God created the Evening Star (Hesperus), giving them matching properties to symbolize their unity of will. Against this background, discovering a match between Hesperus and A_{517} might be compelling evidence for this religion and *against* the Identity Hypothesis. Or suppose there is a respectable and somewhat plausible scientific theory according to which the physical process that generated Hesperus also created a matching sister planet, and this theory specifically predicts that the sister planet is A_{517} . Suppose further that no comparably plausible theory predicts a match between Hesperus and any of the other A s. In a case like this, the discovery that Hesperus matches A_{517} would be evidence for the relevant theory of planetary generation, and against the Identity Hypothesis..

This case resembles the case of psychophysical harmony in important ways. There are many a priori possible correlations between pain and functional properties. Among these, the correlation between pain and F stands out as special in virtue of being uniquely harmonious. For this reason, it's the correlation that is especially likely to hold if a certain not-very-improbable theory is true, namely, that God exists and established the psychophysical correlations. The same cannot be said about the correlation of pain with other functional properties. And, among specific versions of functionalism, there's no obvious reason why the harmony-implying version stands out from the crowd as especially probable.

To sum up: the general observation that there is a correlation between pain and some functional state or other supports functionalism, since functionalism predicts such a correlation. However, this evidence does not support functionalism over the hypothesis that God designed the psychophysical laws to exhibit harmony, since harmony also requires such correlations. On

the other hand, the more specific observation that pain is correlated with F strongly supports the theistic design hypothesis over functionalism, since the former hypothesis predicts that pain will be correlated specifically with F, whereas the latter does not predict that pain will be correlated with F rather than any of the vastly many other states with which it might conceivably be correlated. This more specific observation does support the more specific functionalist hypothesis that pain is F in particular (though not over the hypothesis that God designed the psychophysical laws to exhibit harmony, which also predicts this). But the increase in this more specific hypothesis' predictive power comes with a corresponding decrease to its prior probability, since we had no prior reason to expect this specific functionalist hypothesis to be true rather than any other. But the predictive power of the theistic hypothesis does not come at the cost of an extremely low prior probability, as we argued in §2.4.

3.3 Alternatives to Physicalism and Dualism

We have been treating physicalism as the main alternative to dualism. Let us conclude this section by briefly examining how things look if we consider a somewhat wider range of options for the metaphysics of consciousness. Perhaps the most important alternative to physicalism and dualism is idealism, the view that phenomenal truths are fundamental and physical truths either don't exist or are grounded in phenomenal truths. We think theism should look even *more* attractive if one antecedently takes idealism seriously. It is a familiar point that idealism, at least in its most straightforward Berkeleian form, probably requires (something like) God to explain the order and coherence in our experiences (e.g., the fact that each subject's experience is suggestive of a stable independent environment, and different subjects' experiences are suggestive of the same environment viewed from different perspectives). Similar points hold for Mill's phenomenalism, according to which the physical world is grounded in an array of

phenomenal potentials or “permanent possibilities of sensation.” The most common objection to this view is that, without something like God, the order and coherence in the basic phenomenal potentials involves massive unexplained coincidences (Chalmers 2019, Lee 2016).

Besides idealism, the most important alternative to (traditional forms of) physicalism and dualism is Russellian monism. According to Russellian monism, physics only reveals the relational structure of the physical world, not its intrinsic nature, and the intrinsic qualities (“quiddities”) underlying the relational structures revealed by physics play an essential role in grounding human consciousness. Russellian monism comes in both panpsychist and panprotopsychoist varieties, depending on whether the quiddities are taken to be phenomenal or merely “protophenomenal” properties (i.e., non-structural properties that are not phenomenal, but can ground phenomenal properties when appropriately combined). The Russellian monist agrees with dualism, against traditional forms of physicalism, that consciousness is not grounded in structural truths of the sort revealed by the physical sciences. And Russellian monists commonly agree with dualists that there is a strong link between conceivability and possibility, so that various conceivable scenarios in which the phenomenal truths vary independently of the physical (i.e., structural) truths are genuinely possible. (These are scenarios that would have been realized if the structural roles described by physics had been occupied by different quiddities. For example, one possible quidditative realization of the actual structural truths might have given everyone boring gray phenomenology. Another might have failed to ground macro-level consciousness at all, rendering us all zombies.) Indeed, the ability to uphold a strong link between conceivability and possibility, and to accommodate powerful modal intuitions about the possibility of zombies, inverters, and other phenomenal deviants, is taken to be one of the major advantages of Russellian monism over standard physicalism (Stoljar 2001; Chalmers 2015; Goff

2017). But as Pautz (ms) observes, these features of the Russellian monist view give rise to an explanatory challenge much like that faced by the dualist:

[G]iven the vast (infinite) different combinations of [...] quiddities, and corresponding macro-experiences, why do we have exactly the right combination of quiddities to yield harmonious experiences? Why did we win the “experiential lottery”? This form of the problem [of psychophysical luck] is no less serious than the form of the problem for dualists. (ms: 42)

Given Russellian monism, psychophysical harmony cries out for explanation, and (just as with dualism) theism provides a natural explanation.¹⁷ The argument from psychophysical harmony therefore does not seem to be threatened by bringing in a broader range of options for the metaphysics of consciousness beyond dualism and physicalism.

4. Fine-Tuning and the Multiverse Hypothesis

The argument from psychophysical harmony has some notable similarities with the traditional fine-tuning argument. Both point to certain (presumed) basic features of our universe and claim that these features are significantly more likely on theism than on atheism. The fine-tuning argument relies on facts about fundamental physics and cosmology.¹⁸ Our argument relies on facts about consciousness and the basic laws governing consciousness. Along with many others, we think that the most significant challenge to the traditional fine-tuning argument is the multiverse response, which has been defended by John Leslie (1989), Peter van Inwagen (1993), Derek Parfit (1998), and others. In this section, we explain why multiverse scenarios do not threaten our argument. The argument from psychophysical harmony may therefore enjoy an important advantage over the fine-tuning argument. (We hedge with “may” because we don’t

¹⁷ Pautz notes in passing that “the Russellian monist could offer a theistic explanation” (42).

¹⁸ See Leslie (1989) and Collins (2009) for detailed discussion.

wish to take a stand on whether the multiverse response succeeds in undermining the fine-tuning argument. The issue is extremely complicated.)

The multiverse response involves two key claims. The first is that, given a sufficiently large and varied collection of universes, it's nearly certain that *some* universes will be life-permitting, so the fact that some universe is life-permitting is not surprising, given such a multiverse. Second, since there can only be conscious observers in life-permitting universes, it's unsurprising that we find ourselves in one. The second point invokes the “observation selection effect,” which is essential to the multiverse response. If we have some putative piece of evidence for theism that *isn't* associated with an observation selection effect, multiverse responses fall flat. Given a sufficiently large and varied collection of universes, there are bound to be universes exhibiting all sorts of marvels: stars spelling out messages, “thermodynamic miracles” in which rotting corpses spontaneously come back to life, and so forth. But a naturalistic multiverse hypothesis wouldn't significantly undermine the evidential support these marvels would give to theism, were we to observe them. Even if a naturalistic multiverse makes it likely that such marvels occur somewhere, it doesn't make it likely that we would observe them, since the vast majority of observers in a naturalistic multiverse will find themselves in circumstances where such marvels do not occur. (It is therefore unsurprising that multiverse hypotheses tend not to be invoked in response to theistic arguments from alleged miracles, while such hypotheses are commonly invoked to respond to the fine-tuning argument.)

The key difference between our argument and the fine-tuning argument is that the observation of psychophysical harmony is not strongly associated with an observation selection effect. It is, of course, easy enough to cook up a multiverse hypothesis whose truth would all but guarantee that *some* universe has harmonious psychophysical laws. For example, we might

hypothesize that there is some mechanism that generates a vast number of universes with randomly selected psychophysical laws. Given enough universes, some are bound to achieve psychophysical harmony. But given a naturalistic multiverse of this kind, it should still be deeply surprising that we find ourselves in one of the psychophysically harmonious universes. Now, it may be that *some* harmony in the psychophysical laws is needed for a universe to host anything that counts as a conscious observer at all. Perhaps subjects with nothing but confused, chaotic phenomenology would not count as genuine observers. But there are many ways in which the laws could have deviated from the harmony we find in our universe without undermining our status as genuine observers. For example, there could have been pleasure/pain inversion, or some evaluatively neutral phenomenal state could have played the pleasure role, or there could have been normative disharmony with respect to a smallish subset of our sensory modalities (small enough not to compromise our status as genuine conscious observers). Even if such a naturalistic multiverse could render it unsurprising that our world meets a minimal baseline of psychophysical harmony required for the existence of beings that qualify as conscious observers, it remains surprising that we find ourselves in a universe whose harmony far exceeds that minimum baseline.

There is a rough analogy between the points above and a common response to some specific multiverse hypotheses in the context of the fine-tuning argument. In particular, some naturalistic multiverse hypotheses predict that most observers in existence are Boltzmann brains, short-lived observers that exist within small pockets of localized order, which emerge by chance thermal fluctuations from a vast sea of high-entropy chaos. This is because chance thermal fluctuations will produce small (say, brain-sized) pockets of localized order much more frequently than large (say, planet- or galaxy-sized) regions of order, much as a random text

generator will produce individual English words surrounded by a sea of nonsense much more frequently than it produces whole paragraphs of coherent English. On the assumption that we inhabit such a multiverse, it's extremely surprising that we are not Boltzmann brains and that our surrounding environment is highly ordered. In other words, while it may not be surprising that there is enough order in our vicinity to support the existence of a conscious observer, it's surprising that there is *vastly more* order in our vicinity than this minimum baseline.¹⁹ In our view, any multiverse hypothesis that predicts that the vast majority of observers are Boltzmann brains does very little to undermine the traditional fine-tuning argument. Theism is, we think, better supported by our total evidence than this specific kind of multiverse hypothesis, because theism plausibly makes it much more likely that we would observe not only a life-permitting universe, but a universe where living things exist in a stable form within highly ordered environments, not as short-lived Boltzmann brains.²⁰ (Of course, there may be other reasonable multiverse hypotheses that don't have this implication. For example, Albrecht (2004) argues that inflationary models in cosmology can avoid the problem of Boltzmann brains.) Likewise, theism is better supported by the datum of psychophysical harmony than a multiverse hypothesis which predicts that the vast majority of observers would find themselves in universes much less harmonious than our own.

This isn't to say that psychophysical harmony shows that we don't live in a multiverse. We can cook up a *different* multiverse hypothesis on which most or all universes have harmonious psychophysical laws. But the existence of such a multiverse would *itself* be strong evidence for theism. The multiverse possessing a bias towards harmonious psychophysical laws seems very surprising on standard naturalistic atheism—not much less surprising than our own

¹⁹ Cf. Collins (2009: 256-72).

²⁰ Cf. Collins (2009: 267-8).

universe possessing harmonious laws already was. But if there is a multiverse, it seems much less surprising that it would possess this bias on theism or theism-adjacent views.

5. Conclusion

We've argued that the existence of psychophysical harmony is strong evidence for theism. Psychophysical harmony is much more likely on theism (and on certain theism-adjacent views) than on standard naturalistic atheism, and the prior probability of theism (and, *a fortiori*, the disjunction of theism with these theism-adjacent views) is not so low as to render this uninteresting. Our initial presentation of the argument from psychophysical harmony assumed dualism and the causal completeness of the physical. But as we've seen, standard naturalistic atheism is not significantly helped by rejecting these assumptions and instead assigning a fairly high prior probability to interactionist dualism, physicalism, idealism, or Russellian monism. And while there is a certain analogy between the argument from psychophysical harmony and the more famous argument from cosmological fine-tuning, the most popular objection to the fine-tuning argument does not affect the argument from psychophysical harmony. We conclude that the argument from psychophysical harmony deserves an important place alongside the traditional theistic arguments.

References

- Adams, R. (1992). Flavors, Colors and God. In *Contemporary Perspectives on Religious Epistemology* (pp. 225–240). Oxford University Press.
- Albrecht, A. (2004). Cosmic inflation and the arrow of time. In J. D. Barrow, P. C. W. Davies, & C. L. Harper (Eds.), *Science and ultimate reality: Quantum theory, cosmology and complexity* (pp. 363–401). Cambridge University Press.
- Armstrong, D. M. (1968). *A Materialist Theory of the Mind* (Issue 22). Routledge.
- Aydede, M. (2018). A Contemporary Account of Sensory Pleasure. In L. Shapiro (Ed.), *Pleasure: A History* (pp. 239–266). Oxford University Press.
- Balog, K. (2012). In defense of the phenomenal concept strategy. *Philosophy and Phenomenological Research*, 84(1), 1–23.
- Chalmers, D. (1996). *The Conscious Mind: In Search of a Fundamental Theory*. Oxford University Press.

- Chalmers, D. (2006). Perception and the fall from Eden. In T. S. Gendler & J. Hawthorne (Eds.), *Perceptual Experience* (pp. 49--125). Oxford University Press.
- Chalmers, D. (2015). Panpsychism and panprotopsyism. In *Consciousness in a physical world: Perspectives on Russellian monism*. Oxford University Press.
- Chalmers, D. (2018). The Meta-Problem of Consciousness. *Journal of Consciousness Studies*, 25(9–10), 6–61.
- Chalmers, D. (2019). Idealism and the Mind-Body Problem. In W. Seager (Ed.), *The Routledge Handbook of Panpsychism* (pp. 353–373). Routledge.
- Chalmers, D. (2020). Debunking Arguments for Illusionism about Consciousness. *Journal of Consciousness Studies* 27 (5-6):258-281.
- Collins, R. (2009). The teleological argument: An exploration of the fine-tuning of the universe. In W. L. Craig & J. P. Moreland (Eds.), *The Blackwell Companion to Natural Theology* (pp. 202--281). Blackwell.
- Corabi, J. (2014). The Misuse and Failure of the Evolutionary Argument. *Disputatio*, 6(39), 199–227.
- Cutter, B. (2017). The metaphysical implications of the moral significance of consciousness. *Philosophical Perspectives*, 31(1), 103–130.
- Draper, P. (2017). Atheism and Agnosticism. *Stanford Encyclopedia of Philosophy*. Retrieved Jan. 24, 2021 from <https://plato.stanford.edu/entries/atheism-agnosticism/>.
- Dretske, F. (1995). *Naturalizing the Mind*. MIT Press.
- Enoch, D. (2010). The Epistemological Challenge to Metanormative Realism: How Best to Understand It, and How to Cope with It. *Philosophical Studies* 148(3): 413-438.
- Forrest, P. (2012). Replying to the Anti-God Challenge: A God Without Moral Character Acts Well. *Religious Studies* 48(1): 35-43.
- Goff, P. (2017). *Consciousness and fundamental reality*. Oxford University Press.
- Goff, P. (2018). Conscious Thought and the Cognitive Fine-Tuning Problem. *Philosophical Quarterly*, 68(270), 98–122.
- Griffin, D. R. (2004). *God, Power, and Evil: A Process Theodicy*. Westminster John Knox Press.
- Hawthorne, J., & Isaacs, Y. (2018). Fine-Tuning Fine-Tuning. In M. A. Benton, J. Hawthorne, & D. Rabinowitz (Eds.), *Knowledge, Belief, and God: New Insights in Religious Epistemology* (pp. 136–168). Oxford University Press.
- Heathwood, C. (2007). The reduction of sensory pleasure to desire. *Philosophical Studies*, 133(1), 23–44.
- Hendricks, P. (2018). Sceptical theism and the evil-god challenge. *Religious Studies* 54(4): 549-561.
- Jackson, F. (1982). Epiphenomenal Qualia. *Philosophical Quarterly*, 32(April), 127–136.
- Korman, D. Z., & Locke, D. (forthcoming). Against Minimalist Responses to Moral Debunking Arguments. *Oxford Studies in Metaethics*.
- Langsam, H. (2011). *The Wonder of Consciousness: Understanding the Mind Through Philosophical Reflection*. MIT Press.
- Latham, N. (2000). Chalmers on the addition of consciousness to the physical world. *Philosophical Studies*, 98(1), 71–97.
- Law, S. (2010). The Evil God Challenge. *Religious Studies* 46(3): 353-373.
- Lee, G. (2013). Materialism and the Epistemic Significance of Consciousness. In U. Kriegel (Ed.), *Current Controversies in Philosophy of Mind* (p. 222). Routledge.

- Lee, G. (2016). Worlds, Voyages and Experiences: Commentary on Pelczar's Sensorama. *Analysis*, 76(4), 453–461.
- Leslie, J. (1989). *Universes*. Routledge.
- Leslie, J. (2001). *Infinite Minds: A Philosophical Cosmology*. Oxford University Press.
- Lin, E. (2020). Attitudinal and Phenomenological Theories of Pleasure. *Philosophy and Phenomenological Research*, 100(3), 510–524.
- Locke, J. (1689). *An Essay Concerning Human Understanding* (K. Winkler, Ed.). Hackett.
- Lycan, W. G. (1996). *Consciousness and Experience*. MIT Press.
- Milburn, J. (forthcoming). Evilism and the a priori. *Religious Studies*.
- Miller, C. (2016). Is theism a simple hypothesis? On the simplicity of omni-properties. *Religious Studies* 52(1): 45-61.
- Miller, C. (forthcoming). Why God is probably good: a response to the evil-god challenge. *Religious Studies*.
- Mørch, H. H. (2017). The evolutionary argument for phenomenal powers. *Philosophical Perspectives*, 31(1), 293–316.
- Mørch, H. H. (2020). The Phenomenal Powers View and the Meta-Problem of Consciousness. *Journal of Consciousness Studies*, 27(5–6), 131–142.
- Moreland, J. P. (2008). *Consciousness and the Existence of God: A Theistic Argument* (Issue 1). Routledge.
- Nagel, T. (2012). *Mind and Cosmos*. Oxford University Press.
- Oppy, G. (2006). *Arguing About Gods*. Oxford University Press.
- Page, B. (forthcoming). Arguing to theism from consciousness. *Faith and Philosophy*.
- Page, B. and Baker-Hytch, M. (2020). Meeting the Evil God Challenge. *Pacific Philosophical Quarterly* 101(3): 489-514.
- Papineau, D. (2002). *Thinking about Consciousness*. Oxford University Press.
- Parfit, D. (1984). *Reasons and Persons* (Issue 2, pp. 311–327). Oxford University Press.
- Parfit, D. (1998). Why Anything? Why This? In T. Crane & K. Farkas (Eds.), *Metaphysics: A Guide and Anthology*. Oxford University Press.
- Pautz, A. (ms). *A dilemma for Russellian monists about consciousness*.
<https://philpapers.org/rec/PAUCRM>
- Pautz, A. (forthcoming). How to Achieve the Physicalist Dream Theory of Consciousness: Identity or Grounding? In G. Rabin (Ed.), *Grounding and Consciousness*.
- Pautz, A. (2010). Do theories of consciousness rest on a mistake? *Philosophical Issues*, 20(1), 333–367.
- Pautz, A. (2017). The significance argument for the irreducibility of consciousness. *Philosophical Perspectives*, 31(1), 349–407.
- Pautz, A. (2020). Consciousness and Coincidence: Comments on Chalmers. *Journal of Consciousness Studies*, 5–6, 143–155.
- Poston, T. (Forthcoming). On the Intrinsic Probability of Grand Explanatory Theories. *Faith & Philosophy*.
- Rasmussen, J. and Leon, F. (2019). *Is God the Best Explanation of Things? A Dialogue*. Palgrave MacMillan.
- Rasmussen, J. (2019). How to See the Good in the Foundation. In J. Rasmussen and F. Leon, *Is God the Best Explanation of Things? A Dialogue*. Palgrave MacMillan.
- Robinson, W. (2007). Evolution and epiphenomenalism. *Journal of Consciousness Studies*, 14(11), 27–42.

- Saad, B. (2019). A Teleological Strategy for Solving the Meta-Problem of Consciousness. *Journal of Consciousness Studies*, 26(9–10), 205–216.
- Schaffer, J. (forthcoming). Ground functionalism. *Oxford Studies in Philosophy of Mind*.
- Smart, J. J. C. (1959). Sensations and Brain Processes. *Philosophical Review*, 68(April), 141–156.
- Stoljar, D. (2001). Two Conceptions of the Physical. *Philosophical and Phenomenological Research*, 62(2), 253–281.
- Swinburne, R. (2004). *The Existence of God*. Oxford University Press.
- Swinburne, R. (2009). How the Divine Properties Fit Together: A Reply to Gwizada. *Religious Studies* 45(4): 495-498.
- Tye, M. (1995). *Ten Problems of Consciousness: A Representational Theory of the Phenomenal Mind* (Vol. 282, pp. 606–609). MIT Press.
- Tye, M. (2003). A theory of phenomenal concepts. In A. O’Hear (Ed.), *Royal Institute of Philosophy Supplement* (pp. 91–105). Cambridge University Press.
- Tye, M. (2008). *Consciousness Revisited: Materialism Without Phenomenal Concepts*. MIT Press.
- van Inwagen, P. (1993). *Metaphysics*. Westview Press.
- Ward, K. (2015). The Evil God Challenge--A Response. *Think* 14(40): 43-49.
- Weaver, C. (2015). Evilism, moral rationalism, and reasons internalism. *International Journal for Philosophy of Religion* 77: 3-24.
- White, R. (2005). Explanation as a guide to induction. *Philosophers’ Imprint*, 5, 1–29.
- Wielenberg, E. (2010). On the Evolutionary Debunking of Morality. *Ethics* 120(3): 441-464.
- Yudkowski, E. (2015). *Rationality: From AI to Zombies*. Machine Intelligence Research Institute.